

**RUNNING MEMORY/WORKING MEMORY: SPAN TESTS AND
THEIR PREDICTION OF HIGHER-ORDER COGNITION**

A Thesis
Presented to
The Academic Faculty

by

James M. Broadway, Jr.

In Partial Fulfillment
of the Requirements for the Degree
MASTER OF SCIENCE in the
School of PSYCHOLOGY

Georgia Institute of Technology
MAY 2008

COPYRIGHT 2008 BY JAMES BROADWAY

**RUNNING MEMORY/WORKING MEMORY: SPAN TESTS AND
THEIR PREDICTION OF HIGHER-ORDER COGNITION**

Approved by:

Dr. Randall W. Engle, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Dan Spieler
School of Psychology
Georgia Institute of Technology

Dr. Paul Corballis
School of Psychology
Georgia Institute of Technology

Date Approved: March 28, 2008

This study is dedicated to the antecedents of memory.

ACKNOWLEDGEMENTS

I wish to thank Rob Abraham, Ashley Ahrens, Mike Bunting, Randy Engle, Rich Heitz, Maggie Ilkowska, Tom Redick, and Nash Unsworth for their invaluable support in bringing this investigation into existence. They deserve the bulk of credit for any positive contributions it may make to the field of cognitive psychology, but I must take sole responsibility for any errors introduced herein.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
SUMMARY	ix
<u>CHAPTER</u>	
1 INTRODUCTION	1
Complex and simple memory span	2
Running memory span	6
Exposure	7
Working memory ‘updating’	8
Focus of attention	9
Rate	11
Summary of issues and aims	13
2 METHOD	18
Participants	18
Materials and Apparatus	18
Procedure	18
3 RESULTS	22
Descriptive statistics and zero-order correlations	22
Multiple regressions	24
Interim summary	29
List length effects on memory	31

List length effects on prediction	36
Interim summary	40
3 DISCUSSION	42
Limitations	47
APPENDIX A: Analyses of variance tables	49
REFERENCES	51

LIST OF TABLES

	Page
Table 1: Zero-order correlations	23
Table 2: Means and standard deviations	24
Table A1: Running memory span analysis of variance	49
Table A2: Simple memory span analysis of variance	50
Table A3: Complex memory span analysis of variance	50

LIST OF FIGURES

	Page
Figure 1: Illustration of variance partitioning	25
Figure 2: Partitioning gF variance	28
Figure 3: List length effects on complex and simple span	34
Figure 4: List length effects on running span	35
Figure 5: List length and prediction by complex and simple span	37
Figure 6: List length and prediction by running span	38

SUMMARY

Different versions of complex, simple, and running tests of immediate memory span were compared in their ability to predict fluid intelligence (gF). Conditions across memory tasks differed in terms of whether or not a secondary cognitive task was interleaved between to-be-remembered items (complex versus other span tasks), whether or not more items were presented than were ultimately to-be-remembered (running versus other span tasks), and whether presentation rate was relatively fast or slow (running and simple span tasks). Regressions indicated that up to 42.6% of variance in general fluid gF was explained by the memory span measures entered in different combinations. Across comparisons, shared relationships among span tasks accounted for a plurality of total variance in gF. Results indicate that in spite of procedural differences and resulting intra-individual variance in memory performance, the present memory tasks captured largely the same inter-individual variance in working memory capacity, insofar as this is important for higher-order cognition.

CHAPTER 1

INTRODUCTION

I set out to answer the basic question “does a memory span task need to have a secondary task interleaved between to-be-remembered items in order to be a good measure of working memory capacity?” The question reduces to “do complex span tasks uniquely tap the important dimensions of working memory?” For most of the past three decades, there has been general agreement in the affirmative (see Unsworth & Engle, 2007a,b). However, recent work has suggested that working memory capacity, insofar as it is important for complex cognition, can also be measured well by a simple memory span task (Colom, Rebollo, Abad, & Shih, 2006; Kane, Hambrick, Tuholski, Wilhelm, Payne, & Engle, 2004; Unsworth & Engle, 2007a,b; 2006a). It is clear that the secondary task procedure of complex span tasks affects immediate memory performance. It is plausible, however, that this does not mean that complex span tasks uniquely capture inter-individual variance in such things as simultaneous *storage-and-processing* (Daneman & Carpenter, 1980) or *executive attention* (Engle, 2002). It could be that such abilities are also reflected in performance on memory paradigms lacking this procedure. I used complex and simple span tasks much as others have done before to examine this question (e.g., Engle et al., 1999; Kane et al., 2004; LaPointe & Engle, 1990; Unsworth & Engle, 2007a,b; 2006a). Complex and simple span tasks were used to predict individual differences in general fluid intelligence (gF). I extended this line of research to include a third kind of memory span task, running memory span (Pollack, Johnson, & Knaff, 1959). Running span tasks do not include a secondary task, but like complex span

tasks, running span tasks have been proposed to uniquely tap into executive cognitive functions by virtue of a special procedure, i.e., presenting more items than can be, or are instructed to be, remembered.

I assessed the amount of predicted variance in gF that was due to unique relationships with each span task and the amount that was due to common relationships among the set of span tasks. The question was whether one or more of the span tasks would uniquely predict substantial amounts of gF variance, at the expense of shared prediction. If complex span tasks require executive control processes over and above simple or running span tasks, then they should account for variance in gF over and above. The same may be predicted for the running span task. Alternatively, an experiment might show that what is *common* to the span tasks requires sufficient executive control of working memory, i.e., they all tap the same basic cognitive processes. In this case, variance shared among span tasks would account for the most variance in higher-order cognition.

Complex and simple memory span

Simple span tasks are tests of immediate memory in which a person is exposed to short lists of items and then must report the items in order when the list ends. Complex span tasks include the same serial order memory requirement, but in between presentations of the to-be-remembered items the participant must perform a secondary task such as solving a math equation (Operation Span) or judging a sentence (Reading Span). These so-called *storage-and-processing* tasks were developed to tap the construct of a dynamic working memory system (Baddeley & Hitch, 1974; Daneman & Carpenter,

1980), better than simple storage-only memory tests, thought to reflect a more static short-term memory buffer (Turner & Engle, 1989).

For almost three decades investigators of individual differences in working memory capacity (WMC) have shown that complex span tests are consistently predictive of higher order cognition (for reviews see Ackerman, Beier, & Boyle, 2005; Unsworth and Engle, 2007a). In one influential study, Engle et al. (1999) demonstrated that the variance unique to complex spans (after controlling for simple spans) predicted gF, but the variance shared between the two measures did not. Such findings have stimulated a range of theories of individual differences in WMC, holding in common the view that special and unique properties are conferred on complex span tasks by virtue of including a moderately demanding secondary task in the procedure. Additionally, a great deal of work has been devoted to investigating aspects of the secondary task, such as content domain, its relationship to specialized skills, or its similarity to the to-be-remembered items.

Recent work has called into question the basic assumption underlying much of this work, namely that working memory is uniquely measured by complex span tasks of the storage-and-processing type, and that other memory paradigms tap some other memory system or function altogether, such as *episodic* memory or *short-term* storage. For example, Unsworth and Engle (2007a,b; 2006a) demonstrated that performance on complex and simple spans is affected similarly by experimental manipulations and that simple spans can be just as good at predicting gF (if variance from longer lists is used). Indeed Unsworth and Engle (2007b) have recently shown, in part by reanalyzing Engle et al.'s (1999) data, that in fact it is the variance *shared* between the two span tests that

accounts for the most variance in gF, not variance unique to the complex span tests as hitherto believed (see also Colom, Rebollo, Abad, & Shih, 2006). Additionally, WMC measured by complex span shows close relationships with performance in a wide variety of other memory paradigms, including those that seem to heavily involve retrieval from longer-term memory such as immediate, delayed, and continuous distractor free recall; (Unsworth, 2007; Unsworth and Engle, 2007a) or the letter *fluency* task (Rosen & Engle, 1997). Cantor and Engle (1993) showed that increased response times due to *fan interference* (Anderson, 1974) provided the same information as complex span performance did concerning individual differences in reading comprehension. WMC-related differences in proactive interference have already been noted (Kane & Engle, 2000; Rosen & Engle, 1998).

Unsworth and Engle (2007a,b) have recently proposed a framework for understanding individual differences in WMC that acknowledges separable contributions from two system components: An attention-like *primary memory*, important for maintaining representations over short times, and a *secondary memory*, important for recovering information that is continuously lost from primary memory due to interference from new input to the system. In their review of the relevant literature, Unsworth and Engle (2007a) classified thirteen situations in which individual differences in WMC have been found according to preponderance of demand for either active maintenance in primary memory or controlled search of secondary memory. For example, WMC-related differences have been shown in the *antisaccade* paradigm (Hallett, 1978), where the participant's only job is to look away from, rather than toward, an attention-capturing stimulus (Kane, Bleckley, Conway, & Engle, 2001; Unsworth, Schrock, & Engle, 2005).

This effect would seem to reflect individual differences in primary memory, dedicated to robustly maintaining task goals in WM. However, the antisaccade task would seem to require little in the way of retrieval from secondary memory. A similar interpretation can be applied to the fact that individuals lower in WMC are more susceptible to interference in the Stroop paradigm (Kane & Engle, 2003), where the main role of memory is to maintain the instruction to name the color in which a color-word is printed instead of reading the word (Stroop, 1935). In contrast, WMC-related differences in e.g., a Brown-Peterson paradigm that maximizes proactive interference (Kane & Engle, 2000) could be ascribed mainly to cue-driven search of secondary memory and not so much to active maintenance in primary memory (Unsworth and Engle, 2007a; Table 4, p 123).

Notably, of the several research paradigms examined by Unsworth and Engle (2007a), only three were supposed to display individual differences in WMC reflecting *joint* contributions from primary and secondary memory, namely, immediate free recall and complex and simple span tasks. Their process account suggested that to-be-remembered items in complex span tasks were displaced from maintenance in primary memory by the interleaved processing task. This has the consequence that at time of memory test, cue-dependent search of secondary memory must support responding. In contrast, there is no secondary task in simple span tasks to displace to-be-remembered items from primary memory. Displacement occurs in simple span tasks when the number of to-be-remembered items exceeds the capacity of primary memory (approximately 4 ± 1 items; Cowan, 2001). Therefore, simple span performance on short, ‘sub-span’ lists reflects mostly maintenance in primary memory, but performance on long, ‘supra-span’ lists reflects a combination of maintenance and retrieval from secondary memory. In

contrast, complex span performance reflects a combination of primary memory maintenance and secondary memory retrieval even on very short lists. As part of their support of this account, Unsworth and Engle (2007a; 2006a) showed that the magnitude of correlation between simple span memory and a gF test (Raven, Raven, & Court, 1998) increased as the number of to-be-remembered items increased in the span test. No such pattern was found across complex span list-lengths.

The first purpose of the present investigation was to extend with new data the earlier findings of Unsworth and Engle concerning complex and simple span tests and prediction of higher order cognitive abilities. Two of each kind of span test was used to predict individual differences in gF. Shared and unique contributions to prediction were assessed, and list-length effects on prediction were re-examined for the two kinds of span task. I expected to find that both complex and simple span tests (with optimal administration and scoring) were strongly related to gF, and that shared relationships between span tests would account for most of the gF variance. I did not make predictions as to whether complex span or simple span tasks would correlate more highly with the gF tests, or whether one type of span would account for more criterion variance uniquely than the other. Indeed, results were expected to underscore fundamental underlying commonalities across simple and complex span tasks rather than surface differences between them. The main comparison pertained to the presence or absence of a secondary cognitive task interleaved between to-be-remembered items.

Running memory span

The second purpose of the present investigation was to include the running memory span task (Pollack, Johnson, & Knaff, 1959) in order to widen the search for

commonalities among span tasks. Running memory span tasks are like simple span (and unlike complex span) tasks in that there is no secondary processing task in between to-be-remembered items. However, running span tasks are different from both simple and complex span tasks in that participants report only a subset of items that were presented. For example, a person might be required to report the *last four* items on each trial, but sometimes more than four items (and sometimes exactly four) are presented. In Waugh's (1960) terminology the target set (the portion of the list to be reported at test) is called the *recall series* and the list that is actually presented is called the *exposure series*. When the recall series < the exposure series, such trials can be called *partial recall* trials (Mukunda and Hall, 1992). When the number of to-be-remembered items = the number of presented items, the recall series and exposure series are coincident. Such trials can be called *whole recall* trials and are in fact isomorphic to simple span trials, except that their occurrence within a block is unpredictable.

Exposure

Mukunda and Hall also characterized complex span tests as *partial recall* tests because participants are exposed to information (the interleaved task) that they are not supposed to remember for later test. Relative to simple span tests, this seems the appropriate classification for complex span tests. But relative to running span tests, the classification of complex span tests as whole or partial recall is more ambiguous. In complex span tests, all of the items that are *designated* as to-be-remembered items are, in fact, later tested for memory. In running span tests, items presented are only *potentially* in the target set: Ultimately only some of them are mapped to correct responses at time of memory testing.

The third purpose of the present investigation is to examine individual differences in the effect of being exposed to more items distinguished from the effect of having to remember more items, in relation to performance on gF tests. In whole recall tests like simple or complex span, list length effects on serial order memory confound two effects that are conceptually separable: The effect of having to remember more items and the effect of being exposed to more items. Gates (1916) investigated the *exposure* effect on memory span by presenting items in excess of span. His results showed that the maximum number of items that could be reliably recalled under conventional procedures was much higher than the number that could be recalled when ‘supra-span’ lists were presented, for all but highest-performing ten per cent of individuals. Like Gates’ (1916) procedure, running span tasks are able to dissociate these two effects bound up in list-length effects, by including whole and partial recall trials.

Working memory ‘updating’

In recent years a number of investigators have also argued that partial recall trials in running span can be used to tap a process of WM ‘updating’ (Morris and Jones, 1990; Miyake et al., 2000; Postle, 2003). Within this conceptual framework, WM updating is a kind of executive function that dynamically manipulates short-term memory representations in order to keep the target set current as new items are to-be-remembered and old items are to-be-forgotten (Friedman et al., 2006; Miyake et al., 2000; Morris and Jones, 1990; Postle, 2003). This task-analysis of running span as an ‘updating’ task has some intuitive appeal and fits with increasing interest across disciplines in so-called executive functions. WM updating as measured by the running memory span task has been linked to age-related changes in cognition (Chen & Li, 2007; Fisk and Sharp, 2004; Van der Linden, Brédart, & Beerten, 1994). Friedman, Miyake, and colleagues have used a running memory task as part of an updating latent factor to explain individual

differences across a range of diverse executive functions (Friedman et al., 2006; Miyake et al., 2000), extending this to investigations of genetic heritability (Friedman et al., in press). Running memory span as updating task has been shown to correlate well with complex span tasks (Lehto, 1994; Miyake et al., 2000).

Focus of attention

Cowan and his colleagues have also used running memory span tasks to predict higher order cognition (Cowan et al., 2005) in children and young adults, but without supposing that participants engage and continue an active strategy such as updating. They proposed that participants listen passively to the string of items during presentation, and attention is applied to residual sensory memory traces when the list terminates and the signal to begin recall is given. Trace-decay is forestalled in this manner, and therefore it is possible for more information to be extracted and coded so that a memory can be reported.

The attention-buffer is limited in capacity, imposing a limit on working memory capacity (memory span, ‘span of apprehension’, focus of attention, primary memory, etc.). Cowan (2001) marshaled a wide range of evidence and arguments to support assigning specific number for this limit (4 ± 1 items; Cowan, 2001). Individual differences in the scope and flexibility of attention are reflected in recalling more or fewer items from span memory lists, and are responsible for the demonstrated links between memory span and higher-order cognition. Running memory span is among a somewhat restricted set of tasks endorsed by Cowan as theoretically acceptable measures of WMC or the *scope of attention* (Cowan, 2001; Cowan et al., 2005). Running span in the format used by Cowan and colleagues, like running span in the ‘updating’ format, has

shown strong relationships with criterion cognitive abilities, largely due to variance shared with complex span tasks (Cowan et al., 2005).

There is a range of opinion as to what psychological constructs or processes are measured by running span tasks and consequently there is great diversity in procedures in terms of seemingly important variables such as rate, modality, stimuli, and list structure. Investigators of ‘executive updating’ tend to present items (consonants) visually at rates slow enough to support active rehearsal and updating operations on the target set (e.g., Postle, 2003). Lists are often structured so as to permit a distinction between partial and whole recall trials (or updating and non-updating trials, respectively). That is, sometimes the recall series = the exposure series (i.e., non-update trials, also serving as ‘catch trials’ so that participants do not adopt a strategy of simply ignoring the first item or two on every list). When the recall series < the exposure series (i.e., update trials) it is by increments of one, two, ... k items so as to be able to estimate the number of ‘updating operations’ necessary (Morris and Jones, 1990). Items are presented at either the ‘standard’ rate for span tasks (one item per second), or else slower. Indeed, Postle (2003) argued that for active working memory updating to be assumed, items must be presented no faster than one item every two seconds.

In contrast, investigators of the *scope of attention* present to-be-remembered items at rates presumably too fast to support rehearsal or active manipulation of the target representations during list exposure. This procedure also does not typically include any whole recall (or ‘catch trials’). Additionally, presented lists are not incrementally related in length to the number of items to be reported at test as in the ‘updating’ approach described above. Rather, all presented lists far exceed the number of items that can be

committed to memory after a single exposure, a task feature designed to ‘overload’ WM and prevent participants from adopting a strategy of attempting to memorize the entire list (Cowan, 2001; Cowan et al., 2005).

Rate

The basic contrasts between methods employed by those who conceive of the running memory paradigm as an *updating* or *focus of attention* task coincide with the methodology used by Hockey (1973). Hockey presented auditory digit strings at varying rates (one, two, or three items per second) and manipulated instructions between-subjects. The *passive* group was told to “Listen passively to the list as a whole. Don’t rehearse items or groups of items” (p 106). The *active* group was told to “Concentrate on the items as they arrive, trying to form them into groups of three... rehearse each group of three in turn” (p 106). Hockey’s results showed a crossover interaction such that the active instructions group recalled the most items in the slowest task and the fewest items in the fastest task. In contrast, the passive instructions group remembered the most items at the fastest rates and the fewest items at the slowest rates. The running span procedure typically used by Friedman, Miyake, and colleagues (Letter Memory; 2000; 2007; 2008) correspond closely to the one per second rate and active instructions in Hockey’s (1973) studies, except that items in the former are visual letters instead of auditory digits. Participants in their task are indeed instructed and trained to rehearse incoming items in groups of three (or four, depending on the study). In contrast, the typical running span procedure used by Cowan, Bunting, and colleagues corresponds closely to the passive instructions, fast rate condition in Hockey’s (1973) studies.

A fourth purpose of the present investigation was to examine rate effects on running memory span (as well as simple span) both in terms of memory performance and prediction of higher order cognition. In light of methodological differences between running span measures, it is difficult to say whether the criterion variance that is accounted for by the different running span tests in previous studies reflects processes unique to each, or shared between, the ‘updating’ and ‘focus of attention’ implementations of the task. Bunting, Cowan, & Saults (2006) directly compared performance on fast (2 digits per second) and slow running span tasks (1 digit per second), and found that more items were recalled during the slow than the fast version. Furthermore, serial position functions derived from performance in the slow running span were shallower compared to the fast task, suggesting to Bunting et al. that maintenance-type rehearsal had prevented items from decaying during the slower lists. However, their results shed no light on the question of differential prediction of higher-order abilities as a function of rate because no criterion measures relevant to this question were included in their study. I was motivated to begin to address this knowledge gap in the present work.

I compared performance in a running span task given at a rate that was presumably too fast (two items per second) to support effective maintenance-type rehearsal or active updating strategies, to performance in a running span given at a rate that indeed slow enough (one item per two seconds) to support such processes if participants should choose to attempt them (without being explicitly instructed to do so). If the two running span tasks account for much the same variance in higher-order cognition and shared much the same variance with the gF criteria and other working memory tasks, it would be taken as indirect evidence against the ‘updating’ account of

performance in the running span. The simple span tasks were matched in rate to the two running span tasks as a source of additional information about rate effects, in a task in which no executive updating in the sense described above was hypothesized to occur.

Summary of issues and aims

Current investigators use running span tests in two very different formats to investigate two theoretically distinct WM constructs. However, direct evidence is needed before I can conclude that the different versions in fact measure altogether different psychological processes or system components. Rate effects, modality effects, and exposure effects differ across relevant studies indeterminately. Given that no cognitive task can be regarded as process-pure, it seems plausible instead that performance on fast and slow running spans tap a number of shared processes (as well as some that are unique to each). I addressed this question in the present study by looking at whether fast and slow running span would differentially predict performance on measures of gF.

By the same argument, and just as complex and simple spans are strongly related to each other and predict higher order cognition similarly (due to process overlap, i.e., both tap individual differences in primary capacity *and* retrieval from secondary memory), it is plausible that performance in running span (irrespective of rate, or perhaps differentially by rate) will be strongly related to performance in these two different span measures (again, due to process overlap). Furthermore, it is plausible that a great deal of variance in higher order cognition will be accounted for by variance shared among all three kinds of span task (suggesting that they all tap individual differences in primary memory capacity as well as retrieval from secondary memory).

Based on the work of Unsworth and Engle, I predicted that most of the variance in gF accounted for by the three types of span test would be mostly due to shared relationships among predictors. If including a secondary task in a span test elicits executive cognitive processes additional to those elicited by span tests lacking such a

procedure, then complex span tasks should show unique relationships with higher-order cognition over and above relationships shared with simple and running span tasks. Likewise for running span, with respect to the partial recall procedure as an index of working memory ‘updating’.

Researchers who use running span to investigate ‘updating’ (eg., Friedman et al., 2006) and those who use it to investigate the ‘scope of attention’ (Cowan et al., 2005) have separately found their tasks to correlate strongly with complex spans, simple spans, and higher-order cognition. It has been suggested that the composition of mental processes underlying running memory span performance is sensitive to the rate at which to-be-remembered items are presented. There is asymmetry. The fast rate is argued to limit effective central executive ‘updating’ or rehearsal, so these processes are hypothesized to be absent from running span performance at a fast rate. However, there is no reason to suppose that processes available in the fast task are not also available in the slow task. There are two outcomes to consider.

If the extra cognitive processes allowed by the slow rate of presentation of to-be-remembered items are especially important for doing well on tests of higher-order cognition (e.g., executive functions), then the slow span tasks should show consistent unique relationships with gF over and above relationships shared with fast span tasks. In contrast, if the extra processes allowed by the slow rate are not very important for doing well on gF tests (e.g., maintenance-type rehearsal), then the slow task should not show such unique relationships with gF over and above the fast tasks. Indeed, insofar as maintenance-type rehearsal is argued to ‘contaminate’ a working memory test that allows it (e.g., Cowan, 2001), the slow span tasks might even correlate only weakly with gF and the complex span tasks.

I used multiple regression/correlation approaches to try to answer questions about the common and unique psychological processes tapped into by several individual span tasks. Given that the span tasks all require the same basic response, serial order memory,

it would be reasonable to expect that all will be highly inter-correlated. Here I must outline my approach to deciding how much interpretive weight to give shared and unique variance in this situation. If any of the following outcomes occurred in my results, I would take it to suggest that the predictors are *not equivalent measures* of the same underlying working memory capacity, insofar as this capacity is important for higher-order cognition. Say two span tasks, X1 and X2, are used to predict gF, Y. If X1 and X2 are each correlated with Y but not with each other, then each will account for unique variance in Y but will explain little variance in Y due to shared variance. Such an outcome could be taken as evidence that different cognitive operations are tapped by the span tasks, and these different processes are separately important for gF. If X1 is correlated with X2 but not with Y, it will act as a *suppressor* variable, boosting the correlation between X2 and Y (Cohen, Cohen, West, & Aiken, 2003). Obviously again, little criterion variance will be explained due to variance shared between X1 and X2. The cases where only one predictor is correlated with the criterion and the predictors are not inter-correlated, and where neither task is correlated with the criterion, do not need to be considered. The preceding outcomes would suggest fundamental differences between span tests in terms of underlying cognitive processes elicited by different task variables.

The following outcomes would be taken as evidence suggesting that the span tests are basically equivalent as tests of working memory capacity. Consider the case where X1 and X2 are significantly correlated with each other and with Y. In this case much of the criterion variance explained would be due to shared variance between predictors. But there is always a portion of unique variance in regressions, and one of the predictors may capture more of that than the other one. The problem is how much interpretive weight to give to the unique versus the shared variance. Of logical necessity, one of the two predictors will be more highly correlated (numerically) with the criterion than the other one is. Since the predictors are highly inter-correlated it can occur that the predictor with the higher correlation with Y will uniquely account for criterion variance over and above

the shared prediction, and subsume the contribution from the predictor with the lower correlation with Y.

In this case it is largely a matter of judgment (after considering other available evidence) to decide whether to emphasize in one's interpretation the shared relationships among predictors, and between these and the criterion, or the unique relationships between one predictor and criterion. It makes sense to look for a stable pattern across comparisons in deciding this issue. If unique relationships between a given span task and gF are consistent across comparisons, it might suggest something interesting going on. If the unique portions of variance explained by a given span task were changeable across comparisons, it would be like chasing a ghost to interpret the unique variance associated with that task. Whatever is most stable across comparisons should be interpreted.

In the case where predictors are significantly inter-correlated and significantly correlated with criterion variables, it makes more sense to focus on the variance shared among all the tasks, since this will be likely to reflect common underlying processes. Occasions of unique prediction by one or other task could be due to any number of reasons meaningful, statistical, or accidental, and unless a stable pattern emerges they should be given less interpretive weight than the shared predictive variance.

The remaining text is organized as follows. I describe the method of the experiment, then report and discuss individual differences results at a *macro* level. Specifically, I assess global indices of memory span as predictors of individual differences in gF. After showing that each of the span measures has good predictive validity, I show that this result is mostly due to variance shared among the present span tests. I conclude that a shared WMC construct supports performance across the various span tests. Next, analyses on a more *micro* level are reported. Specifically, I examine effects on memory performance of the experimental variables list-length and rate of presentation in the simple and running span tests, and list-length in the complex span tests.

The point of the task-level analyses is mainly to support or provide information concerning the individual differences findings. Examining the effects of list-length and other task variables can provide information concerning underlying processes that might or might not be at work across the span tasks. The results will show that the tasks “worked as expected” given available knowledge concerning serial order memory tasks and the effects of rate, number of to-be-remembered items, number of presented items, or interleaved secondary task. I finish by examining list-length effects on prediction of gF in the manner of Unsworth and Engle (2006a). The final section of this paper discusses the present findings and their implications for theories of working memory capacity and its measurement.

CHAPTER 2

METHOD

Participants

Ninety-four participants between the ages of 18 and 35 years of age (mean = 23.57, SD = 4.38) were recruited from the Atlanta community and were individually tested in a sound-attenuated booth after informed consent was obtained, in exchange for financial compensation.

Apparatus and Materials

All tests were programmed in E-prime experimental software (Schneider, Eschman, & Zuccolotto, 2002) and presented on a personal computer. In the memory span tasks participants viewed sequences of black capital letters (F, H, J, K, L, N, P, Q, R, S, T, or Y) presented one at a time in 28 point bold Arial font in the center of the screen against a gray background.

Procedure

In Session 1, participants performed two complex span tasks (Operation and Reading Span), as well as a set of Raven's Standard Progressive Matrices (Raven, Raven, & Court, 1998). Each of the complex span tasks is a storage-and-processing dual-task procedure requiring mental operations interleaved with item encoding and recall. In Operation Span, participants solved simple math equations interleaved with the presentation of individual letters for later recall. After 3- 7 trials (randomly determined) a cue prompted participants to report in serial order the letters they had been shown. There were three trials for each list-length. Participants responded by clicking on the cells of a 4 x 3 grid displaying the twelve letters of the pool of items from which to-be-

remembered items were sampled across trials (Unsworth, Heitz, Schrock, & Engle, 2005). Participants were instructed to click a 'Blank' button for any items they could not remember, and to click 'Clear' if they wanted to begin over their response sequence. Participants clicked a 'Next' button to end the response period and proceed to the next trial. Reading Span was identical to Operation Span, except that concurrently with remembering letters, participants read sentences and indicated whether or not they made sense instead of solving math equations. In scoring the span tasks one point was given for each item correctly selected in correct serial order, regardless of whether or not the entire set was perfectly recalled (Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005).

The set of Raven's Matrices in the present study consisted of twelve spatial reasoning problems. Each problem presented a rectangular matrix of geometric figures with a missing element. Participants selected from an array of choices at the bottom of the screen the figure that would complete the overall pattern of the matrix. Participants had five minutes to complete as many problems out of twelve as they were able and one point was assigned to each correct answer. Task order in Session 1 was the same for all participants: Operation Span, Reading Span, and Raven's. Session 1 lasted approximately one hour.

Participants returned to the lab on a separate day for Session 2, during which they performed fast and slow simple and running spans to test WMC, and Shipley's Abstraction Series to test gF. In Session 2 all participants first performed the two simple span tests, then Shipley's Abstractions, followed by the two running span tests. The rate order was counterbalanced across participants, such that half performed tasks the faster span test first, followed by the slower one (separately for simple and running span tests). The other participants were given the span tests in the reverse rate order. Participants were assigned to one of these two conditions in the counterbalanced order in which they

arrived for testing. Session 2 lasted for approximately one hour. In Session 1 and Session 2, participants worked without an experimenter present in the testing booth and were monitored for compliance by means of a closed-circuit camera system.

Session 2 tasks are described next in the general order in which they were administered. The simple span tests used the same pool of twelve letters as the complex span tests in Session 1 and the response format was also identical. Participants saw lists of 3- 9 items to be reported in the same order in which they were shown. There were three trials for each list-length. Letters appeared on the screen for 300 ms in both fast and slow versions. Stimuli followed each other by 200 ms in the faster one and 1700 ms in the slower one, yielding a rate of presentation of approximately two letters per second in the former test and one letter per two seconds in the latter. Scoring was identical to that for the complex span tests.

Shipley Abstractions Series consisted of twenty incomplete alphanumeric series presented individually on the screen one after the other. Participants were required to type in the letter(s), number(s), or word that would complete the series. For example, if shown “mist-is wasp-as pint-in tone-__” the correct answer would be “on.” Participants had five minutes to complete as many problems as they were able and one point was assigned for each correct answer.

The running span tests were closely matched in procedure and materials to the simple and complex span tests. The same pool of 12 letters was used and the response format was identical. Participants were required to recall the most recent 3- 8 letters during one block of trials each, in random order. If the number of to-be-remembered items equaled n , the number of presented items equaled $n + 0$, $n + 1$, $n + 2$, and $n + 3$, in

random order. The lists were blocked by recall series and participants were informed at the start of a block of trials how many items from the end of a list they would need to remember. Within a block, participants saw three lists in which the number of to-be-remembered items equaled the number of presented items ($n + 0$), and one list each where they differed by one item ($n + 1$), by two items ($n + 2$), and by three items ($n + 3$). The preponderance of $n + 0$ trials ('catch' trials) was motivated by the wish to strongly discourage participants from ignoring the first item or two from each list. Technically, the maximum score in each of the running span tests was 198. In the multiple correlation/regression analyses, I restricted the running span data to include trials where the number of to-be-remembered items < the number of presented items.

At the start of each block of trials, a screen instructed the participant how many of the last letters they should try to remember for that block. This screen remained visible until the participant clicked the mouse to proceed to viewing the letter strings. Participants were given no instructions whatsoever about adopting an active or passive strategy, just as in the complex and simple span tests. Rate of presentation was manipulated in the same manner as in the simple span tests. Participants responded just as in the complex and simple span tests, i.e., by selecting items from a grid that presented all the possible letters that could appear on a trial. Participants were informed again on the response screen how many of the last letters they should try to report in correct order. Scoring was just as in the other span tests.

CHAPTER 3

RESULTS

Multiple regression/correlation results are reported in this section. Results show that all of the tests were significantly inter-correlated, and that a stable proportion of variance in gF was predicted by variance that shared among the span measures. More detailed task-level analyses, e.g., the effects on performance due to varying the number of to-be-remembered items, number of shown items, rate of presentation, etc., are reported afterwards. Effects on prediction of gF by list length in the span tasks are also examined.

Descriptive statistics and zero-order correlations

Running span trials in which the number of to-be-remembered items = the number of presented items, i.e., so-called ‘catch trials’ or *whole recall* trials, are formally similar to simple span trials. Therefore the analyses reported here exclude such running span data in order to limit multicollinearity with other tests. This means that multiple regression/correlation results reported in this section reflect performance on running span trials where the number of to-be-remembered items < the number of presented items, i.e., so-called ‘update trials’. This selection of data has the benefit of isolating the task variable that most distinguishes running span from complex or simple span tests. The excluded data are recovered in later sections reporting more detailed micro-level results.

Zero-order correlations among the tests are presented in Table 1, below. Means and standard deviations are reported in Table 2. The running span scores reflect three trials each of trying to remember 3- 8 letters in order, and simple span scores three trials attempting to remember 3 – 9 letters in order. The complex span results reflect memory performance on three trials at list-lengths 3-7. Cronbach’s alpha estimates of internal consistency are reported on the diagonal of the correlation matrix. Reliabilities are all

adequate to excellent and inter-correlations among all the tests are all significant at $p < .001$. These results establish the existence of strong relationships among all of the span tests and the gF tests. The following section further examines specific components of variance in gF that can be accounted for by the different span tests, individually and in combinations.

Table 1. Zero-order correlations among tests of immediate memory span and gF (N = 94). Raven = Raven's Matrices; Ship = Shipley Abstraction Series; Run 500 and Run 2000 = Running Span at presentation rate of one item per 500 ms and 2000 ms, respectively; O Span = Operation Span; R Span = Reading Span; Simp 500 and Simp 2000 = Simple Span at presentation rate of one item per 500 ms and 2000 ms, respectively. Cronbach's alpha estimates of reliability are on the diagonal. All entries are significant at $p < .001$.

N = 94	Raven	Ship	Run 500	Run 2000	O Span	R Span	Simp 500	Simp 2000
Raven	(.661)							
Ship	.614	(.840)						
Run 500	.494	.615	(.865)					
Run 2000	.444	.624	.776	(.886)				
O Span	.341	.372	.480	.496	(.881)			
R Span	.460	.475	.602	.610	.787	(.895)		
Simp 500	.289	.433	.694	.688	.516	.503	(.765)	
Simp 2000	.533	.536	.739	.709	.605	.675	.715	(.793)

Table 2. Means, standard deviations and ranges. In the span tests one point was assigned for each item correctly recalled in correct serial position. Running span scores reflect only trials where number of to-be-remembered items < number of presented items. In the gF tests one point was assigned for each correctly solved problem.

	Mean (SD)	Range
Raven	7.8404 (2.2012)	10 (2 – 12), max 12
Ship	14.1383 (3.05717)	18 (2 – 20), max 20
Run 500	33.8404 (16.8901)	79 (1 – 80), max 99
Run 2000	43.7979 (18.89733)	88 (4 – 92), max 99
O Span	52.7340 (17.07931)	71 (4 – 75), max 75
R Span	48.7128 (17.85523)	73 (2 – 75), max 75
Simple 500	81.500 (16.39483)	76 (46 – 122), max 126
Simple 2000	90.8723 (17.73899)	100 (22 – 122), max 126

Multiple regressions (variance partitioning)

The high inter-correlations shown in Table 1 justify forming composite variables in the following analyses. Using regressions to analyze criterion variance accounted for after controlling inter-correlation among predictors was described in detail in Chuah and Maybery (1999) and was also used by Cowan et al. (2005) in a study that included both running and complex span test data. A Venn diagram can be used to represent the shared and unique portions of variance accounted for, illustrated in Figure 1, below.

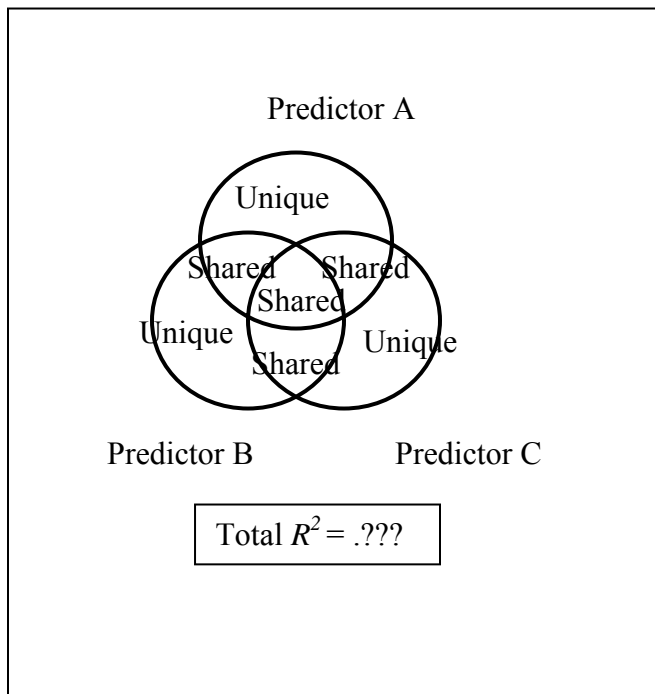


Figure 1. Illustration of variance partitioning using three predictors. The circles represent the predictor variables. Adding the numbers in the sections will give the total amount of criterion variance accounted for (R^2). The numbers that would be entered in the non-overlapping sections represent the proportion of variance accounted for by *shared* relationships between or among predictors, and numbers in non-overlapping sections give the proportion of variance accounted for by variance *unique* to respective predictors.

First I made a criterion variable, gF, by averaging z-scores on Raven's Matrices and Shipley's Abstractions. Then I made composite predictor variables by similarly averaging z-scores on the two complex, simple, and running span tests. Because all inter-correlations were significant, there is no point in reporting significance tests for each of the following sets of regressions. The reader can safely trust that the following sets of regressions all accounted for significant variance in the criterion. Thus, in the following analyses I focus on the questions of *how much* variance in gF is explained, how much is

due to variance shared among predictors, and how much is due to unique relationships between predictors and criterion. Predictors that accounted for significant variance due to unique relationships with the criterion are indicated in Figure 2.

All the span tests together explained about 42.6% of variance in higher-order cognition. Variance partitioning results are presented in Figure 2, Panel A, below. These results indicate that roughly half of the predicted variance was due to shared relationships among the span tests (accounting for 20.4% of criterion variance). The running span tests accounted for the largest portion of variance due to unique relationships with the criterion (accounting for 10.6% of criterion variance). The simple and complex span composites made basically zero unique prediction, deriving their predictive utility in this set of regressions from variance shared with the running span tests. Running and simple span tests also shared a discernible portion of predictive variance (accounting for 8.5% of criterion variance). This is somewhat noteworthy because, as described earlier, the present data excluded the whole recall running span trials (where, like simple or complex span trials, the number of to-be-remembered items equaled the number of presented items).

Decomposing these predictors into their constituent tasks, thus examining them pair-wise, yielded consistent results. In each case, about half of criterion variance explained was due to shared relationships between span tests. Panel B of Figure 2 shows that roughly 41.4% of variance in gF was accounted for by the two running span tests together, and a full three-fourths of this result was due to variance shared between the predictors (accounting for 31.2% of criterion variance). Neither the fast nor the slow

running span contributed more than a small in comparison with the shared portion (explaining 6.8% and 3.4%, respectively).

Panel C of Figure 2 shows that roughly 35.5% of variance in gF was accounted for by the two simple span tests together, and about half of this result was due to variance shared between the predictors (accounting for 16.1% of criterion variance). Here the slower span test made the larger_unique contribution, slightly larger in amount to the shared portion (accounting for 19.3% of criterion variance), while the faster simple span test made basically zero unique prediction. Panel D of Figure 2 shows the results using Operation and Reading Span as predictors. The two complex span tests together accounted for roughly 27.1% of variance in gF, about half of this due to shared relationships between them (accounting for 15.7% of criterion variance). Reading Span made the larger unique contribution in this case (accounting for 11.4% of criterion variance), while Operation Span made zero unique prediction.

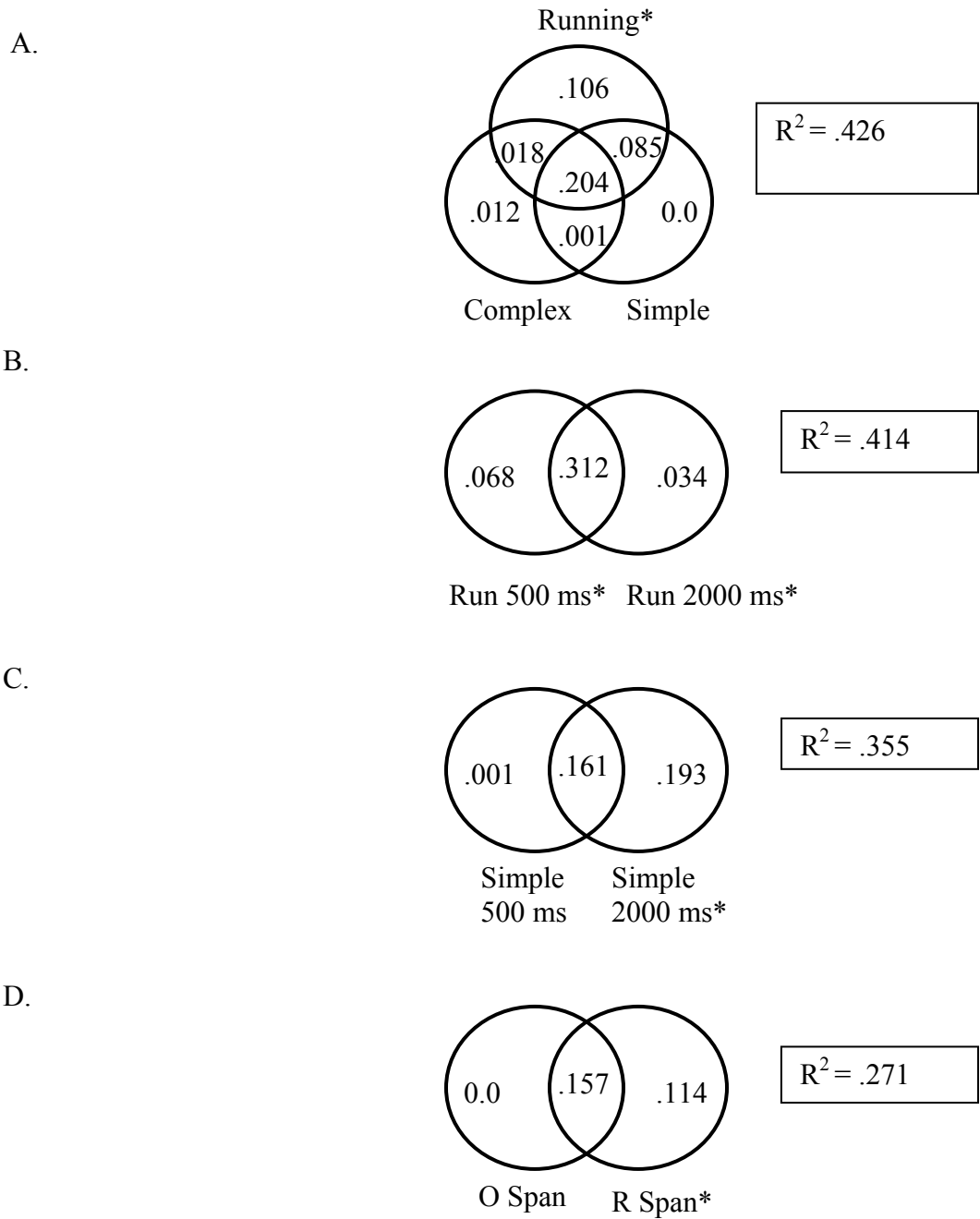


Figure 2. Partitioning gF variance according to shared and unique contributions among predictors. (*) denotes significant unique prediction at $p < .05$.

Interim summary

The preceding results indicate that each of the tests of immediate memory span used in the study reliably predicted individual differences in higher-order cognition. When span tests were used in different combinations, at least half of variance explained was reliably due to variance *shared* among predictors. The consistency of this result underlies my claim that the three types of span tests measure the same underlying working memory capacity, in spite of wide procedural differences. The differential unique prediction by one or the other span test in certain cases should be weighed somewhat lightly against this consistent shared portion, at least until a consistent pattern emerges.

The results present rather strong evidence against the notion that complex span tasks of the processing and storage type, i.e., with an interleaved secondary task, are uniquely valid or inherently superior measures of working memory capacity. In the present study, two kinds of serial order memory test, each lacking a secondary task, proved to be as highly correlated with gF as the complex span tests. Both simple and running span tests, due to variance shared between them, accounted for criterion variance over and above that explained by variance they shared with complex span tests. With respect to the running span vis-à-vis the other span tasks, variance unique to the effect of presenting more items than were to-be-remembered appeared to predict criterion variance over and above that explained by shared relationships with the whole recall tasks, complex and simple spans.

The finding that fast and slow running span accounted for much the same variance in higher-order cognition is not consistent with the idea that active updating processes were driving mean performance differences shown between fast and slow running span. Presenting items at a rate of two per second would seem to preclude effective updating during list presentation, while presenting items at a rate of one every two seconds would seem to fully support updating if participants were to avail themselves of this strategy. One would expect that a drastically different use of strategies across two tasks would lead to drastically different relationships with gF, yet 75% of the variance in gF explained by fast and slow running spans was due to variance shared between them.

The finding that fast and slow running span accounted for much the same variance in higher-order cognition is not consistent with the idea that rehearsal-based processes were driving the mean performance differences between fast and slow running span. Presenting items at a rate of two per second would seem to preclude effective maintenance-type rehearsal, while presenting items at a rate of one every two seconds would seem to fully support rehearsal if participants were to avail themselves of this strategy. Yet it has been argued that increased use by participants of simple strategies like rehearsal will generally decrease correlations of a task with measures of higher-order cognition, and some evidence has been shown to that effect (Engle, Cantor, & Carullo, 1992). Again, drastic differences in strategy across two tasks should result in different relationships with gF, yet 75% of the variance in gF explained by fast and slow running spans was due to variance shared between them. By this same argument, the finding that the slow and fast simple spans shared approximately half of their predictive variance, and

that the slower simple span was if anything more correlated with gF than was the fast simple span conflicts with the idea that maintenance-type rehearsal was driving the performance advantage in the slow versus fast simple spans.

List-length effects on memory performance

I examined list-lengths 3-8 for the simple and running span tests, and list-lengths 3- 7 for the complex span tests in repeated-measures ANOVAs to assess effects of experimental variables on proportion correct recall. Results are summarized in Tables A1, A2, and A3, in the Appendix. The ANOVA results confirm that proportion correctly recalled from each list was strongly affected by experimental variables such as number of to-be-remembered items, number of presented items, or rate of presentation. Due to the number of tasks included, for pictorial clarity I represent recall functions distributed across Figures 3 and 4. Figure 3 shows functions for the two complex span tests and the two simple span tests. The running span data are presented together in Figure 4. There is a separate line for each trial type, i.e., $n + 0$, $n + 1$, $n + 2$, and $n + 3$ lists. Altogether, the present data show an especially thorough sampling of the space of memory performance.

Figure 3 shows that simple span memory was higher than complex span memory across list-lengths (i.e., number of to-be-remembered items). Memory for items in simple and complex span memory was similarly affected by increasing numbers of to-be-remembered items, showing about a 10% increase in forgetting for each additional item. Rate interacted with number of to-be-remembered items in the simple span tasks such that decrements in memory due to the, relative to the slow, rate of presentation were most evident for list-lengths greater than five to-be-remembered items. The equivalent performance between fast and slow simple spans for list-lengths less than about five

items, together with somewhat diverging performance for longer list lengths suggests that rate of presentation does not have much effect within the range of primary memory capacity, but is more influential on memory for lists long enough to require retrieval from secondary memory. These results indicate that if rehearsal-based processes were important for the mean performance advantage for the slow versus the fast simple span, these rehearsal processes were selectively effective at differentiating performance when the number of to-be-remembered items was greater than about five items. Note that the sizes of the rate effect and its interaction with list-length are relatively small in comparison with the main effect of increasing number of to-be-remembered items (Table A1).

Figure 4 shows that increasing number of to-be-remembered items also had a large effect on memory in the running span tasks. The functions for fast and slow running span trials, where the number of presented items equaled the number of presented items, i.e., whole recall or catch trials, are comparable in slope and height to each other, and in slope with the functions for the other span tasks. The basic list-length effects of greater forgetting with increasing number of to-be-remembered items appear to be similar in size across span tasks (Table A2 and Table A3). The running span trials where the number of to-be-remembered items was less than the number of presented items, i.e., partial recall or update trials, show the exposure effect and interactions with rate and with number of to-be-remembered items. The main effect of exposure was nonlinear: A greater increase in forgetting occurred for $n + 1$ trials relative to $n + 0$ trials, than occurred for $n + 2$ relative to $n + 1$ trials, or for $n + 3$ relative to $n + 2$ trials. The exposure effect was greater when the rate was fast relative to when the rate was slow.

The effect of increasing number of to-be-remembered items interacted with the effect of increasing number of presented items. Greater forgetting occurred due to presenting more items than were to-be-remembered, for longer relative to shorter to-be-remembered series.

Rate did not interact with the effect of increasing number of to-be-remembered items like in the small interaction shown in the simple span data, but did interact with the effect of showing more items than were to-be-remembered. This indicates that if rehearsal-based processes were important for the mean performance advantage for the slow versus fast running spans, these rehearsal processes were selectively effective at differentiating performance when more items were presented than were to-be-remembered. Note that like for the simple span data, the main effect of rate was smaller in comparison to the other main effects on running memory span, and the interactions involving rate produced very small effects indeed. Note also that the effect of rate in the simple span data is comparable in size to the effect of secondary task in the complex span data, i.e., the effect of Operation versus Reading Span.

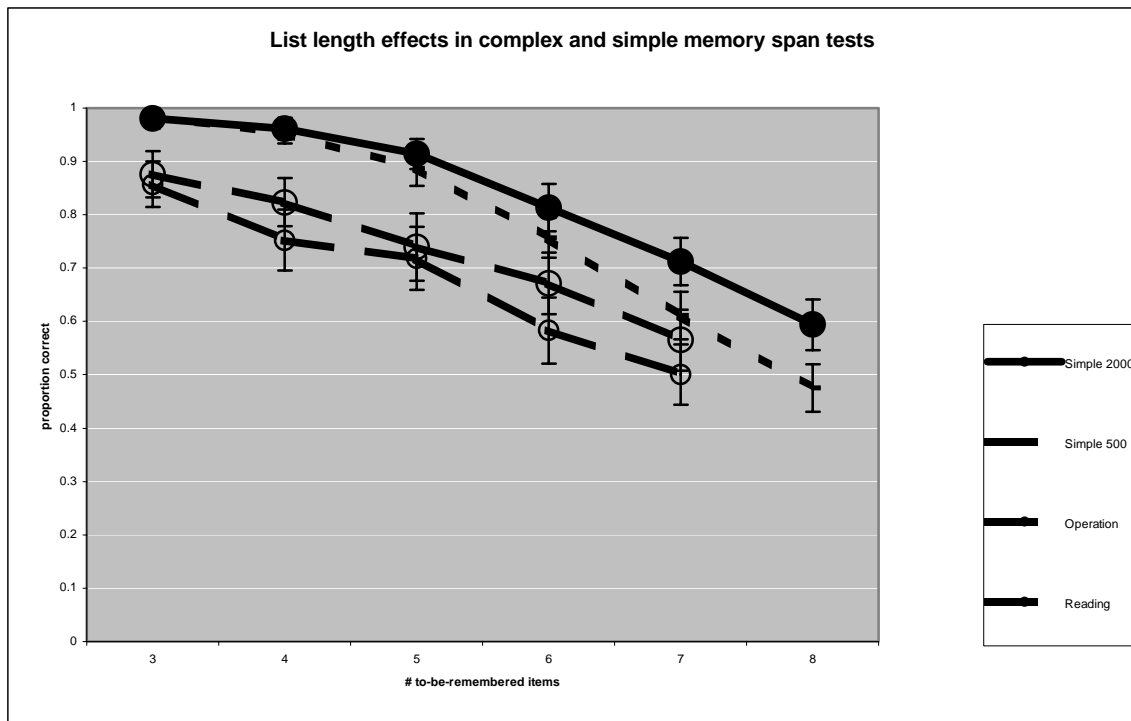


Figure 3. List length effects on memory in simple span (2000 ms), simple span (500 ms), Operation Span, and Reading Span. Error bars depict within-subject 95% confidence intervals.

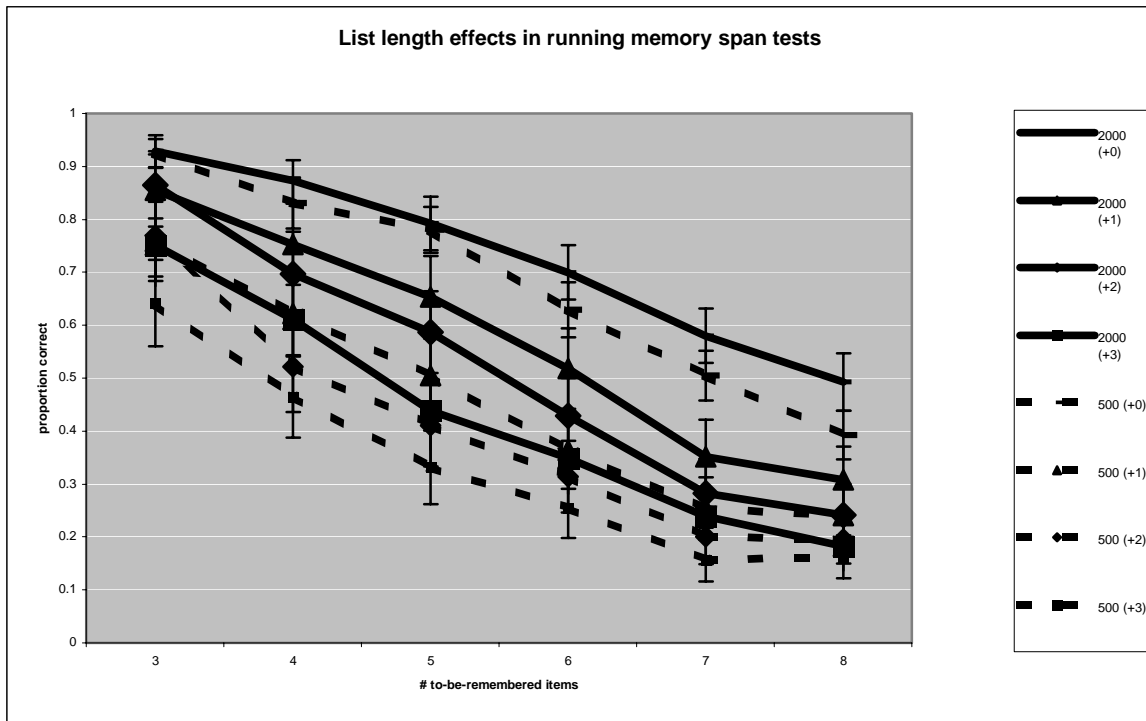


Figure 4. List length effects on memory in two running span tests (2000 ms; = solid lines; 500 ms = dashed lines). Separate functions for $n + 0$, $n + 1$, $n + 2$, and $n + 3$ trial types are shown. Error bars depict within-subject 95% confidence intervals.

List-length effects on prediction of higher-order cognition

Unsworth and Engle (2007a; 2006a) demonstrated that correlations between simple memory span and gF increased with increasing number of to-be-remembered items, but found no such effect for complex memory span. I expected to replicate their positive finding for simple span lists and the null finding for complex span lists. Z-score averages of trials at each list length made up the predictor variables (separately for complex and simple span lists). A test of heterogeneity among correlated correlations (Meng, Rosenthal, & Rubin, 1992) was not significant for the simple span lists, $\chi^2_{(4)} = 1.5878, p > .05$, indicating that in the present sample, remembering a string of three items in correct order was just as predictive of higher-order cognition as remembering seven items in order was. This outcome raises some questions about details of Unsworth's and Engle's (2007ab; 2006ab) account of processes underlying performance in complex and simple span tasks, insofar as their account is based on their positive finding in this regard. I did however, "replicate" their null finding with respect to complex span lists, $\chi^2_{(4)} = 1.877, p > .05$. These results are plotted in Figure 5, below.

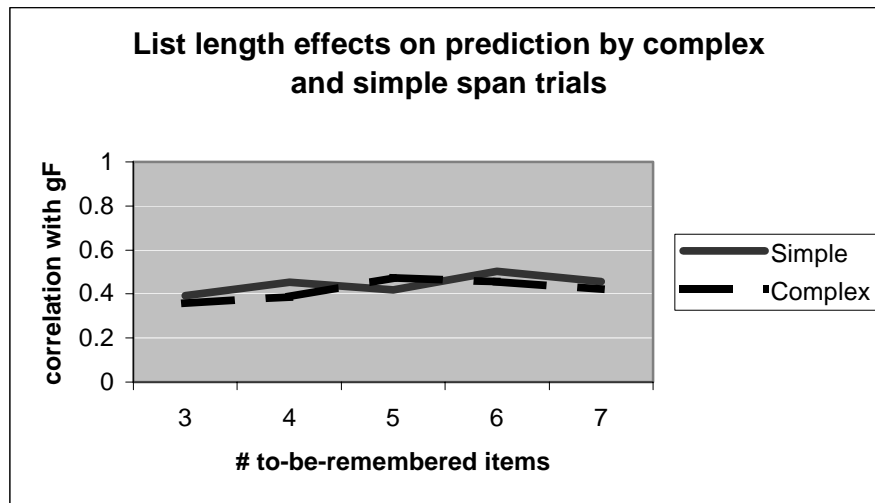


Figure 5. List length effects on prediction by complex and simple span trials. All correlations (and inter-correlations) are significant at $p < .001$

Before extending this sort of analysis to the running span data, consider that the concept of “list-length” in running span trials is fractionated: The number of to-be-remembered items is not identical the number of presented items. Would these conceptually distinct variables show differences in predicting higher-order cognition? I chose to address this question by comparing data from trials that could be matched to another trial in terms of either the number of to-be-remembered items or the number of presented items. For example on trials represented as 3 (4) on the horizontal axis in Figure 6, participants had to remember three items but were exposed to four. On trials represented as 4(4), participants were exposed to the same number of items as in the 3(4) trials, but were required to remember one more. On 4(5) trials, participants tried to remember the same number of items as in the 4(4) trials, but were exposed to one more, etc. Z-score averages of all the running span trials at each list-length (separating number

of to-be-remembered items from number of shown items) formed the composite predictor variables. The resulting “saw-tooth” function showed significant heterogeneity among correlations to gF, $\chi^2_{(9)} = 20.23, p < .01$. These results are plotted in Figure 6, below.

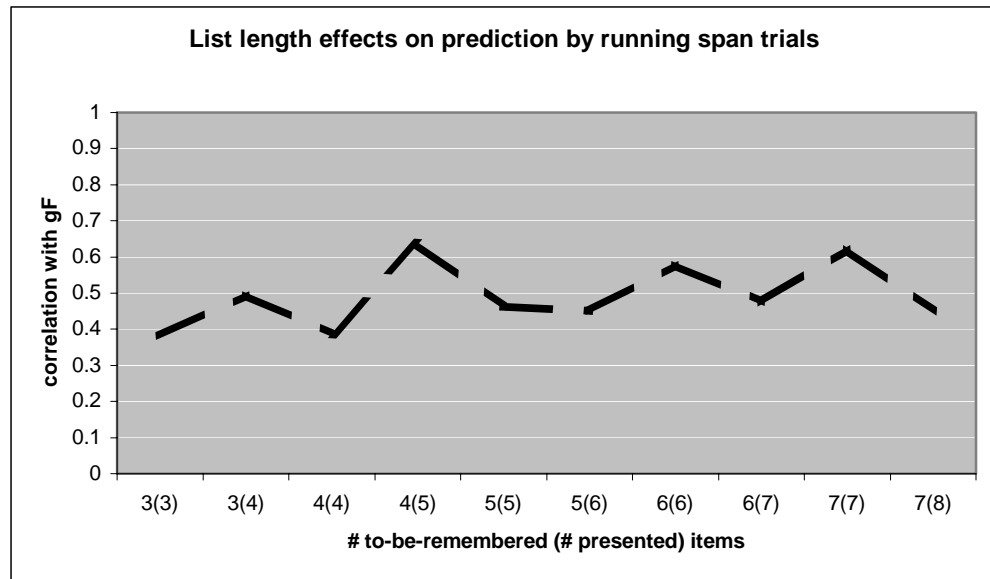


Figure 6. List length effects on prediction by running span trials. All correlations (and inter-correlations) are significant at $p < .001$.

The function suggests that remembering three or four items when four or five, respectively, were shown, was more strongly associated with gF than was remembering three or four items when three or four, respectively, were shown. A single contrast pitting correlations at 3(4) and 4(5) against 3(3) and 4(4) confirmed this impression, $z = 2.6525, p < .01$, 95% CI for the difference = $.4849 \pm .3583$. The function suggests a predictive gain for *partial recall* trials relative to *whole recall* trials, when the number of

to-be-remembered items is around or below the limits of “span”, i.e., within postulated capacity limits of a primary memory or focus of attention (Cowan, 2001; 4 items \pm 1).

The above pattern of differential prediction reversed when target lists exceeded “span.” A single contrast pitting correlations at 6(6) and 7(7) against 6(7) and 7(8) was significant, $z = 2.1148$, $p < .05$, 95% CI for the difference = $.3866 \pm .3858$. Note that the running span tests here included three times as many *whole recall* as *partial recall* observations, and the confidence interval just barely excludes zero. The shape of the function for “supra-span” lists might reflect differences in reliability of measurement, but this argument would seem to work against obtaining the “sub-span” result.

It is possible that the boost in prediction for partial recall, sub-span, running span resulted from participants dropping from ceiling performance at those points. This suggestion would have to also apply to the simple span data however, where at least equivalent ceiling performance was achieved for the shortest lists, yet the ascending function of correlations between simple span list-lengths and gF was not found in the present study, in a failure to replicate Unsworth and Engle’s (2006) results.

It should be emphasized that *all* of the points in Figures 5 and 6 represent significant inter-correlations and correlations with gF at $p < .001$. Consistent with my approach to interpreting the macro-level correlations, overall the analyses of correlations at the list-length level suggest most strongly that the same basic dimensions of working memory capacity are equivalently tapped into, no matter how many items are to-be-remembered or how many are presented.

Interim summary

A wide range of performance was sampled by the several span tasks in the present study. Main effects and interactions produced an orderly family of recall functions. Performance was affected in predictable ways by presenting items at different rates, increasing the number of to-be-remembered items, increasing the number of items shown in excess of the number to-be-remembered, or requiring participants to perform an attention-demanding secondary task in between to-be-remembered items. The wide intra-individual variation in performance across task variables is notable in light of the macro-level results indicating that the tasks accounted for much the same inter-individual variation in gF. These findings overall suggest that much the same underlying working memory construct is tapped by the widely different serial order memory tasks in the present study.

The correlation between gF and complex span did not vary as a function of number of to-be-remembered items, consistent with Unsworth's and Engle's (2006a) demonstration. But neither did the correlation between gF and simple span performance, an outcome that is not consistent with their results. In the present data, remembering a simple string of three items in correct order was just as strongly predictive of higher-order cognition as was remembering a string of seven items in correct order. It should be noted that the apparent discrepancy across studies could be due to the fact that the response format of the present simple span task was not paper-and-pencil, like those in Unsworth's and Engle's (2007a; 200a) analyses. Even if serial order scoring is applied, paper-and-pencil formats cannot control the order in which people write down their

responses. For all span tasks in the present study, response order was constrained to begin with the first target item and proceed in serial order to the last.

The correlation between gF and running span (unlike complex or simple span) varied as a function of to-be-remembered items, but in a joint relation with the number of presented items. Partial recall trials seemed to be more highly correlated with gF than were whole recall trials, when the number of to-be-remembered items was less or equal to about $4 \pm$ items. Whole recall trials seemed to be more highly correlated with gF than were partial recall trials when the number of to-be-remembered items was $>$ about 4 ± 1 items. Note that just like for simple span, remembering a string of three items in correct items in order was significantly correlated with gF. But there was an apparent boost in predicted variance if four items were shown, although only three had to-be-remembered. Etc.

CHAPTER 4

DISCUSSION

The present study addressed a few different issues. First in light of recent publications and re-analyses, I sought to demonstrate again whether complex and simple span tasks, as measures of working memory capacity, account for much the same individual differences in higher-order cognition. They did. When memory is the dependent measure, there is little additional benefit to prediction of individual differences in gF by interleaving secondary tasks between to-be-remembered items. If it were true that complex span tasks tap into executive functions or a dynamic storage-and-processing ability in addition to basic serial order memory ability, one would expect these additional processes to be reflected in unique relationships with higher-order cognition. Such unique relationships between complex span performance and gF were definitely not shown in the present results. Both of the other kinds of span test explained more variance in gF than the complex span tests did.

Note that while the running span, like complex span, has been labeled an executive function task, the simple span test has not been so labeled. If anything, simple span has been argued to tap into a passive storage buffer with little direct connection to cognitive control systems. Yet the simple span tests outperformed the complex span tests at predicting higher-order cognition in the present sample (in fact, particularly so when presentation rate was slow enough to allow rehearsal). These findings suggest that it might be good for working memory researchers to look at what is basically common to span tasks, e.g., serial order memory, rather than become absorbed by what may ultimately be minor differences among them.

Second I extended this approach to include a different serial order memory task, running memory span, and evaluate this task against the others as a measure of working memory capacity that predicts individual differences in higher-order cognition. This project was highly successful. The running span tasks in the present study accounted for about 41.4% of the variance in gF. The running span tasks shared about half of their predictive variance with the complex and simple span tasks, strongly indicating that fundamentally common underlying memory abilities are tapped by all three kinds of span task.

Third I looked at list length effects on memory performance and its prediction of gF in the three kinds of span task. These too were most consistent with the idea that fundamentally common dimensions of working memory capacity are involved in performing all the tasks. In spite of wide variation in performance due to task variables such as the number of to-be-remembered items, the rank ordering of individuals remained consistent across those variables. There were orderly decrements in memory performance for each additional to-be-remembered item, but this did not translate into differential prediction by list length in any clear way. Remembering a simple string of three letters (without interference from a secondary task) was just as highly correlated with gF as performance in many of the more complicated situations sampled in the present design. This suggests that the same basic dimensions of working memory capacity required to remember a short list of three items are required to remember 1) the longer lists in simple span tasks, and 2) the items in complex and running span tasks, short or long lists. This also suggests that these same basic dimensions of working memory capacity are important for higher-order cognition.

Fourth I looked at rate of presentation in the running span tasks, both in terms of memory performance and prediction of higher-order cognition. This was stimulated in part by a split in the literature using the running span task, i.e., updating versus focus of attention. In the present study, items in the running and simple span tasks were presented

at a rate of two per second or one every two seconds. The faster rate was supposed to be fast enough to preclude either effective maintenance-type rehearsal or active concurrent updating strategies. The slower rate was supposed to be slow enough to fully support such strategies if participants chose to avail themselves of them. Even though the faster rate resulted in more forgetting in the running span task, the fast and slow tasks accounted for basically identical individual differences in higher-order cognition. This is not what would be expected if participants spontaneously engaged in an active updating strategy as envisaged by those accepting the executive function account of running memory span, i.e., deleting old items, appending new items, re-ordering items, etc., concurrently with list presentation. Instead these results are consistent with the idea that much the same encoding and retrieval processes are at work across serial memory tasks at various rates of presentation.

Participants may be induced by experimenter instructions to rehearse items in changing groups of three, four, etc. items in order to mimic a kind of active updating of the target set, as in Hockey (1973) and some more recent studies (Friedman et al, 2006; in press; Miyake et al., 2000). It might be informative if data from such shadowing performance were actually reported (or recorded) for studies in which participants are instructed to rehearse/update in such a manner. The extent to which shadowing errors correspond to subsequent recall is potentially informative. Note that rehearsal in running memory span, of pre-target items that are ultimately *not* to-be-remembered, is likely to be detrimental to overall performance, because this would elevate the probability of incorrectly recalling items that had appeared prior to the target items. Furthermore, I suggest that the active strategy instructions procedure is somewhat artificial as an operational definition of whatever working memory updating must be *really* like, although such interventions can lead to potentially useful data about grouping effects in immediate memory (as in the early experimental work with the running memory task, e.g., Hockey, 1973; Pollack, Johnson, & Knaff, 1959).

The problem of *updating in memory in general* was introduced (Bjork, 1978) in terms of changing relatively stable representations such as the name of one's *current* wife, where you parked the car *today*, etc. It seems to me that the postulated executive function of *working memory* updating is not so much wrong as it is redundant with the transient nature of working memory representations as they are already universally conceived. It seems that the notion of working memory updating has been somewhat more usefully approached through specifying biological mechanisms that could serve as gates controlling what currently active information is kept the same and what is changed, dynamically in response to changing environments and goals (e.g., Hazy, Frank, and O'Reilly, 2006). This non-homuncular kind of updating could take place within-lists in lag tasks like running memory span or *n*-back, but also across lists in more conventional simple or complex span tasks. It might also be supposed that the information needed to perform the secondary task in complex spans must be discarded just like previous-list items, consistent with the idea that working memory processes, such as updating, must operate generally across a wide variety of situations. Working memory tasks cannot be neatly classified as 'updating' or 'not updating' tasks. Updating in a very general sense must occur after each list, working against the effects of proactive interference, in any memory task with repeated trials with a limited pool of items. Running span trials cannot be neatly classified as 'update' and 'non-update' trials, as suggested by the executive updating account. This view is based in part on my consistent finding in the present study of substantial predictive variance shared among all the span tasks, whether or not they included special procedures designed to elicit *updating*, *storage-and-processing*, or *passive maintenance*.

I argue that people will use whatever is available at encoding and retrieval in order to remember events from the recent past. Stuff that might be available to help remember would include person variables like internally generated retrieval cues or strategies. Other available supports for memory would include environmental variables

pertaining to the encoding, such as features and contexts defining events when they occur, or external retrieval cues when they are to-be-remembered. Given that early-presented items in running memory span sometimes are and sometimes are not part of the set of to-be-remembered items, an optimal strategy might indeed be to encode all incoming items rather indifferently with respect to whether or not they are to be included in the target set, and to make this discrimination at time of retrieval.

The same basic information needed to perform the running span task in this ‘passive encoding’ manner would be available under either fast or slow task conditions. Many forms of additional information could be available at time of test under slow conditions due to additional encoding, consolidation, or re-coding processes. Features of memory representations might be relatively harder to discriminate under rapid versus slow presentation schedules. Still, while there are innumerable conceivable sources for the advantage for memory performance in the slow running span task relative to the fast one, such processes did not result in the two tasks accounting for much different variance in gF (quite the contrary). Thus I suggest that processes/representations available in *either* fast or slow presentation conditions were primarily driving performance and prediction of higher-order cognition, not processes hypothesized to be available in slow but not in fast presentation conditions.

Finally on the topic of rate and memory, consider that rate is *only relatively* fast or slow in any situation. Human memory must accommodate the fact that events occur in the world at a wide variety of rates, relative to the observer, in limitless combinations along the continuum from fast to slow. While there may be specialized processes or systems that come into play when dealing with special aspects of such *fast-slow* information, at some point accurate memory depends on the integration, or common scaling of such information.

My remarks have led me again and again to conclude that fundamental commonalities among the span tasks in the present study were the decisive influences on

memory performance as well as prediction of individual differences in higher-order cognition. After a thorough consideration of the ways in which the tasks differ, let us consider at last what it is they have most in common. The fact that all the memory tests used in the present study are serial order memory tasks is so basic that it almost escapes notice, and its importance has been possibly under-rated. Consider that in many situations, remembering *when* events occurred can be as crucial as remembering *what* events occurred (e.g., remembering whether one has turned the ignition, *then* put the car into gear, before pressing the gas). Many goal-directed activities require that actions be performed in sequences that are constrained somehow. It would be a good use for a memory if it contained records of what actions had been already performed, and in what order, what occurred next, and so on. This is plausibly the case when one is trying to solve inductive or sequential reasoning puzzles like Raven's Matrices or Shipley's Abstractions.

In keeping with my interest in memory processes that might possess generality across laboratory paradigms, I do not suggest that serial order tasks uniquely tap into order information or temporal organization in memory, although they were explicitly considered as such by Mukunda and Hall (1992). I do suggest however that they can function like a microscope lens of a particular magnification aimed at such phenomena. Such glimpses provided by serial order span tasks into how memory records and supplies information about the temporal order of events can be usefully combined with other research paradigms that offer a view at a different magnification.

Limitations

My conclusions are limited to the specific task parameters of the several span tests. Only two rates of presentation were employed, and these were not all that far apart depending on the frame of reference (two items per second versus one item every two seconds). My results could depend in some way on the choice of rates. My results

cannot be generalized to running memory span tests given in the auditory modality either. The present study represents the first part of an extended parametric manipulation of the running memory span task that will include a wider sampling of rates (one item per 250 ms, one per 1000 ms, and one per 2500 ms) and modalities (auditory and visual). Our conclusions are limited to the two specific measures of higher-order cognition used in the present work. It is likely that a wider sampling of intelligence or behavioral tasks in future investigations would produce greater differentiation among the span tasks.

APPENDIX

Table A1. Summary of results from repeated-measures ANOVA on proportion correct recall across running span task conditions. *Rate* = effect of 500 ms vs 2000 ms between items. *Recall* = effect of the number of to-be-remembered items. *Exposure* = effect of number of presented items.

Source	df	MS	<i>F</i>	<i>p</i>	partial eta squared
Rate	1	9.915	96.747	< .001	.513
Error	92	.102			
Recall	5	32.212	305.109	< .001	.768
Error	460	.106			
Exposure	3	20.995	221.289	< .001	.706
Error	276	.095			
Rate by Recall	5	.114	1.850	= .102	.020
Error	460	.062			
Rate by Exposure	3	.262	3.521	= .016	.037
Error	276	.074			
Recall by Exposure	15	.228	3.200	< .001	.034
Error	1380	.071			
Rate by Recall by Exposure	15	.073	1.373	= .152	.015
Error	1380	.053			

Table A2. Summary of repeated measures ANOVAs on proportion correct across simple span task conditions. *Rate* = effect of 500 ms vs 2000 ms between items. *Recall* = effect of the number of to-be-remembered items.

Source	df	MS	<i>F</i>	<i>p</i>	partial eta squared
Rate	1	.752	21.065	< .001	.186
Error	92	.036			
Recall	5	5.940	257.741	< .001	.737
Error	460	.023			
Rate by Recall	5	.117	8.723	< .001	.087
Error	460	.013			

Table A3. Summary of repeated measures ANOVAs on proportion correct across complex span task conditions. *2nd Task* represents the difference between Operation and Reading Span tasks.

Source	df	MS	<i>F</i>	<i>p</i>	partial eta squared
2nd Task	1	.653	13.908	< .001	.130
Error	93	.047			
Recall	4	3.244	103.984	< .001	.528
Error	372	.031			
2nd Task by Recall	4	.045	1.855	= .118	.020
Error	372	.024			

REFERENCES

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, 131, 30 – 60.
- Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, 3, 288 – 318.
- Baddeley, A.D. & Hitch, G.J. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47 – 89). New York: Academic Press.
- Bjork, R. A. (1978). The updating of human memory. In G. H. Bower (Ed.), *The psychology of learning and motivation*. (Vol. 12., pp. 235-259). New York: Academic Press.
- Bunting, M., Cowan, N., & Saults, J.S. (2006). How does running memory span work? *The Quarterly Journal of Experimental Psychology*, 59, 1691 – 1700.
- Cantor, J. & Engle, R. W. (1993). Working memory capacity as long-term memory activation: An individual differences approach. *Journal of Experimental Psychology: Learning Memory, and Cognition*, 19, 1101 – 1114.
- Chen, T., & Li, D. (2007). The roles of working memory updating and processing speed in mediating age-related differences in fluid intelligence. *Aging, Neuropsychology, and Cognition*, 14, 631 – 646.
- Chuah, Y.M.L., & Maybery, M.Y. (1999). Verbal and spatial short-term memory: Common sources of developmental change?. *Journal of Experimental Child Psychology*, 73, 7 – 44.
- Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Colom, R., Rebollo, I., Abad, F.J., & Shih, P.C. (2006). Complex span tasks, simple span tasks, and cognitive abilities: A reanalysis of key studies. *Memory & Cognition*, 34, 158 – 17.
- Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O., & Engle, R.W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769 – 786.

- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87 – 185.
- Cowan, N., Elliott, E.M., Saults, J.S., Morey, C.C., Mattox, S., Hismjatullina, A., & Conway, A.R.A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, 51, 42 – 100.
- Daneman, M. & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, 19, 459 – 466.
- Engle, R.W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11, 19 – 23.
- Engle, R.W., Cantor, J., & Carullo, J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 972 – 992.
- Engle, R.W., Tuholski, S.W., Laughlin, J.E., & Conway, A.R.A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128, 309 – 311.
- Fisk, J.E. & Sharp, C.A. (2004). Age-related impairment in executive functioning: Updating, inhibition, shifting, and access. *Journal of Clinical and Experimental Psychology*, 7, 874 – 890.
- Friedman, N.P., Miyake, A., Corley, R.P., Young, S. E., Defries, J.C., & Hewitt, J.K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, 17, 172 – 179.
- Friedman, N.P., Miyake, A., Young, S.E., DeFries, J.C., Corley, R.P. & Hewitt, J.K. (in press). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*.
- Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, 18, 1279 – 1296.
- Hazy, T.E., Frank, M.J. & O'Reilly, R.C. (2006). Banishing the homunculus: Making working memory work. *Neuroscience*, 139, 105--118.
- Hockey, R. (1973). Rate of presentation in running memory and direct manipulation of input-processing strategies, *Quarterly Journal of Experimental Psychology*, 25, 104 – 111.
- Gates, A.I. (1916). The mnemonic span for visual and auditory digits. *Journal of Experimental Psychology*, 1, 393 – 403.

- Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General*, 130, 169 – 183.
- Kane, M.J. and Engle, R.W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 336 – 358.
- Kane, M. J. & Engle, R. W. (2003). Working memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132, 47 – 70.
- Kane, M.J., Hambrick, D.Z., Tuholski, S.W., Wilhelm, O., Payne, T.W., & Engle, R.W. (2004). The generality of working memory capacity: A latent variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133, 189 – 217.
- LaPointe, L.B. & Engle, R.W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 1118 – 1133.
- Lehto, J. (1996). Are executive function tests dependent on working memory capacity?. *The Quarterly Journal of Experimental Psychology*, 49A, 29 – 50.
- Meng, X., Rosenthal, R., & Rubin, D.B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111, 172 – 175.
- Miyake, A. Friedman, N. P., Emerson, M.J., Witzki, A., Howerter, A., & Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49 – 100.
- Morris, N. & Jones, D. (1990). Memory updating in working memory: The Role of the central executive. *British Journal of Psychology*, 81, 111 – 121.
- Mukunda, K, & Hall, V.C. (1992). Does memory for order correlate with performance on standardized measures of ability? A meta-analysis. *Intelligence*, 16, 81 – 97.
- Pollack, I., Johnson, L.B., & Knaff, P.R. (1959). Running memory span. *Journal of Experimental Psychology*, 57, 137 – 146.
- Postle, B.R. (2003). Context in verbal short-term memory. *Memory & Cognition*, 31, 1198 – 1207.
- Raven, J.C., Raven, J.E., & Court, J.H. (1998). *Progressive Matrices*. Oxford, England: Oxford Psychologists Press.

- Rosen, V. M. & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language*, 39, 418 – 436.
- Rosen, V. M. & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, 126, 211 – 227.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-prime user's guide. Pittsburgh: Psychology Software Tools Inc.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: General*, 121, 15-23.
- Turner, M.L. & Engle, R.W. (1989). Is working memory task-dependent? *Journal of Memory and Language*, 28 127 – 154.
- Unsworth, N. (2007). Individual differences in working memory capacity and episodic retrieval: Examining the dynamics of delayed and continuous distractor free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 1020 – 1034.
- Unsworth, N. & Engle, R.W. (2006a). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, 54, 68 – 80.
- Unsworth, N. & Engle, R.W. (2006b). A temporal-contextual retrieval account of complex span: An analysis of errors. *Journal of Memory and Language*, 54, 346 – 362.
- Unsworth, N. & Engle, R.W. (2007a). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114, 104 – 132.
- Unsworth, N. & Engle, R.W. (2007b). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, 133, 1038 – 1066.
- Unsworth, N., Heitz, R.P., Schrock, J.C., and Engle, R.W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37, 498 – 505.
- Van der Linden, M., Brédart, S., & Beerten, A. (1994). Age-related differences in updating working memory. *British Journal of Psychology*, 85, 145 – 152.
- Waugh, N. (1960). Serial position and the memory- span. *American Journal of Psychology*, 73, 68 – 79.

Zachary, R. A. (1986). *Shipley Institute of Living Scale: Revised manual*. Los Angeles: Western Psychological Services.